Le Centre canadien de recherche sur l'informatisation du travail

Canadian Workplace Automation Research Centre

CWARC's Research and Development Program in Machine-Aided Translation

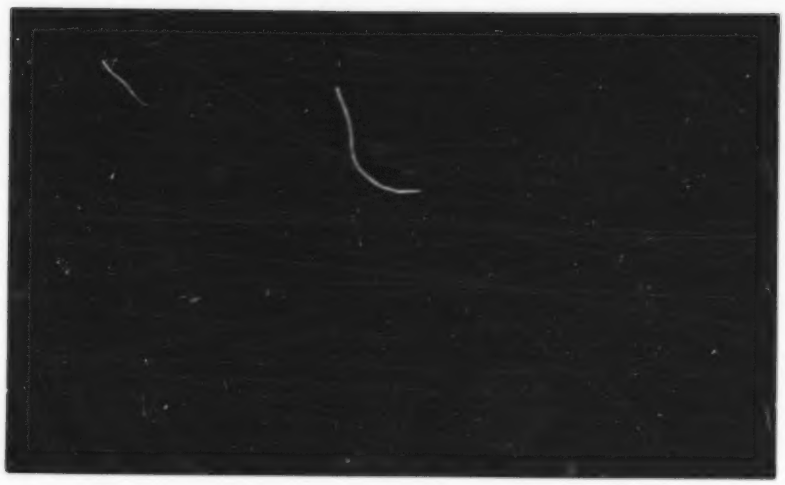
by

Pierre Isabelle
&
Elliott Macklovitch

Researchers

Canadian Workplace Automation
Research Centre

Canada



CWARC's Research and Development Program in Machine-Aided Translation

by

Pierre Isabelle
&
Elliott Macklovitch

Researchers

Canadian Workplace Automation
Research Centre

French version available

CO28-1/21-1988E ISBN 0662-16412-1

It is generally agreed that fully automatic high-quality translation of unrestricted texts (FAHQTUT, as it is sometimes called) is an unattainable goal, at least for the foreseeable future. This is because the proper translation of natural language texts routinely requires extra-linguistic knowledge, i.e., information that is not explicitly conveyed by the words of the text alone. While substantial progress has been made over the last thirty years in getting computers to extract the linguistic meaning of a sentence, we still know relatively little about how to represent, store and access vast amounts of real world knowledge or how to get a machine to emulate common sense reasoning in ways that would enable it to translate even as well as a novice translator.

From this, it follows that if computers are to make a significant contribution in meeting the ever-growing worldwide demand for translation, the ultimate goal of FAHQTUT must be scaled down and sub-optimal goals established for machine-aided translation. Compromises are possible along three distinct axes:

. the degree of automation;
. the range of texts a system is intended to handle;
. the quality of the translations it will produce.

Of these, the last is perhaps the least interesting: few people appear willing to accept approximate translations containing unpredictable errors of the sort a machine makes. Alternatively, one can develop systems that are less than fully automatic, either by leaving the initiative for translation with the human translator and providing him with a set of tools to perform associated tasks such as terminological research or multilingual text processing more efficiently - one then speaks of machine-aided human translation; or by transferring the initiative for translation to the computer, with the human providing assistance on problems that are too difficult for the machine to resolve - this is known as human-aided machine translation.

To date, all general purpose machine translation systems have required some degree of human intervention, either before, during of after the actual

translation process. In most cases, the weight of this interaction has been either too heavy for translators to tolerate or too great for the systems to be cost effective.

The third suboptimization technique attempts to rectify this situation by trading generality for quality. Systems developed under this approach are not intended to translate arbitrary text but are specifically designed for restricted sublanguages: either naturally occurring sublanguages, such as weather bulletins or stock market reports; or stipulated subsets of natural language, such as the varieties of customized English that impose linguistic restrictions on those who draft documentation in an effort to make the texts more easily comprehensible to men or machines. The most successful application of the sublanguage approach to MT is the well-known METEO system, which the federal Translation Bureau operates to translate over 20 thousand words of weather bulletins a day.

The program of CWARC's R&D team in the machine-aided translation is based on the appreciation of the current state of the art outlined above. The group's long-term research goal is to study ways of making MT systems more intelligent, i.e., less literally bound to the wording of the input text, by adapting techniques that have proven successful in other domains of artificial intelligence.

Particular attention will be paid to the notion of translation rule in the next generation of MT systems and to the question of how contrastive knowledge of language pairs is best represented and acquired. As a mid-term objective, the group hopes to identify one or more promising sublanguages for which dedicated MT systems on the METEO model could be developed. And finally, a project is currently underway to design a specialized translator's workstation, which will ergonomically integrate a set of computerized tools that should allow today's translator to more efficiently perform routine administrative, research and text manipulation tasks and thus become more productive. If you have any questions, please contact the authors.